Thai Localization

Theppitak Karoonboonyanan
`thep@linux.thai.net`

Thai Linux Working Group

6 September 2011

Section 1

Demonstration

- Input Method & Rendering
- Editing
- Word Break
- Firefox
- Office Suite
- Other Applications
- Translations

Section 2

Technical Details

- XIM
    - Fixed X11R6 Xlib source (via XFree86, now in X.Org)
    - 3 XKB layouts (Ketmanee, Pattachote, TIS-820.2538)
    - Added read/write **StringConversionCallback** support (i.e. context-sensitive sequence check/correction)
- GTK+
    - Context-sensitive IM API in GTK+ 2
    - Context-sensitive correcting IM with **im-xim**
    - Client-side supports
    - Thai-Lao GTK+ IM Module in upstream GTK+
    - **gtk-im-libthai** – LibThai-based plug-in

- SCIM
  - LibThai-based **scim-thai**
    - 3 layouts (Ketmanee, Pattachote, TIS-820.2538)
    - 3 strictness levels (Passthrough, Basic, Strict)
    - Surrounding text support, with correction
  - Other engines:
    - **scim-tables-additionals** (no check)
    - **scim-m17n** (surrounding text support + correction)
- iBus
  - Current engines:
    - **ibus-table-thai** (no check)
    - **ibus-m17n** (preedit)
  - Surrounding text API requested (implemented)
  - LibThai-based engine not done yet

- Pango
  - WTT-based clustering
  - PUA Substitution-based shaping for legacy Win & Mac fonts
  - OpenType-based glyph positioning
- Mozilla
  - Pango-based
- Qt 4
  - OpenType-based glyph positioning
  - No support for legacy Win& Mac fonts
- LibreOffice
  - ICU-based shaping
  - Graphite support

- Utilities
  - **swath**: Thai word break filter for LaTeX, HTML, RTF
- Libraries
  - **libthai**: basic routines for Thai language processing, including dictionary-based word segmentation
  - **ICU**: with Thai word segmentation support
- Pango
  - **pango-libthai**: libthai-based language engine for Thai word break
  - now incorporated upstream

- Mozilla
  - Pango-based
- Qt 4
  - `dlopen()` **libthai** for word break
- LibreOffice
  - Based on ICU

- GNU C library
  - All ISO/IEC 14652 locale categories
- CLDR
  - Available

- THAILATEX:
  - babel-based Thai support for LATEX
- console-data
  - fonts & keymap available in classic console
  - new console, `console-setup`, shares fonts & keymaps with X

- GNOME
  - Started around 2003 (GNOME 2.4)
  - 2 active volunteers + occasionnal contributors from Ubuntu
  - Current: $\approx 80\%$ translated
- KDE
  - Started around 1998
  - 2 active volunteers
  - Current: $\approx 42\%$ translated

- National Fonts (NECTEC)
  - Recommendations for Thai font creation
  - 3 standard, free families (2 serif, 1 san-serif)
- UI Font (NECTEC)
  - Font with specialized metrics for screen display
- TLWG
  - Monospace fonts
  - Hand-written font
  - DejaVu-compatible display font
  - Decorative font contributed from f0nt.com
- DIP-SIPA
  - 10 contest-winning fonts of new designs
  - License problem solved, under releasing process

- Linux TLE (NECTEC)
  - Ubuntu-based
  - Targets Thai desktop users
  - Additional free software from NECTEC research labs
- Suriyan/ThaiOS (SIPA)
  - Ubuntu-based
  - Targets Thai desktop users
  - Additional free software from SIPA promotion projects
- KKLUG Live (Khon Kaen Linux User Group)
  - Debian-based
  - Targets Thai desktop users
  - Mainly Debian testing snapshots + patched packages from DebianClub